ROPES &GRAY

1301 K STREET, NW, SUITE 800 EAST WASHINGTON, DC 20005-7008 202-626-3900 F 202-626-3961
BOSTON NEW YORK SAN FRANCISCO WASHINGTON, DC

March 31, 2003

Mark A. Greenwood

Ms. Rebecca Kane
U.S. Environmental Protection Agency
Office of Enforcement and Compliance Assurance
Mail Code 2222A
1200 Pennsylvania Avenue NW
Washington, DC 20460

Dear Ms. Kane:

On behalf of the Coalition for Effective Environmental Information, I am submitting the attached comments on EPA's Enforcement & Compliance History Online Website. We appreciate the opportunity to provide input on this new information product. If you have any questions, please feel free to contact us.

Best regards,

Mark A. Greenwood

7169929_1

Before the United States
Environmental Protection Agency

Enforcement & Compliance History On-line Project

Comments of the Coalition for Effective Environmental Information

---

On November 20, 2002, the U.S. Environmental Protection Agency (EPA) posted on its Website the "Pilot Phase" version of an information product called Enforcement & Compliance History On-line (ECHO). This product is a database that seeks to portray compliance data and other information for approximately 800,000 facilities that are regulated by EPA.[1] In its announcement of ECHO, EPA has asked for general comments on the site as well as specific comments on the accuracy of data about particular facilities.

The Coalition for Effective Environmental Information (CEEI) appreciates this opportunity to provide general comments on the ECHO site. CEEI is not offering comments on the data presented about specific facilities, reserving that task to its individual members. CEEI is a group of major companies and business organizations, representing a wide array of industry sectors, that share a common interest in improving how government collects, manages, uses and disseminates environmental information.[2] CEEI supports public policies that encourage data quality, governmental accountability, efficient data collection, alignment of data with strategic goals and consistent management of environmental information resources.

CEEI sees ECHO as a highly precedential information product, perhaps the most important Website established by EPA under the Bush Administration. The site raises major questions of information policy in areas of interest to CEEI. The quality of the data in the ECHO system and the appropriate presentation of those data are significant questions to be addressed. The site is, therefore, an important case study in how EPA is implementing the Information Quality Guidelines recently issued both by the Agency and the Office of Management and Budget.

Since ECHO is attempting to present sensitive, and potentially controversial, information about approximately 800,000 facilities throughout the country, the site also presents challenges for EPA and the states to assure that they remain accountable to both data submitters and users who are expecting information that is accurate, understandable and environmentally relevant. ECHO also represents a significant experiment in linking its publicly stated goals for reduction

---

[1] ECHO presents compliance information for EPA programs under the Clean Air Act (CAA), Clean Water Act (CWA) and Resource Conservation and Recovery Act (RCRA).

[2] CEEI includes representatives from the aerospace, chemical, energy, automobile, pharmaceutical, forest products, petroleum, electronics and consumer products industries.

of public risk to compliance measures. In particular, EPA has linked compliance information about each facility with demographic data about the community surrounding that facility, a step that EPA has never taken before in any of its major Websites on environmental releases from regulated facilities.

CEEI appreciates that EPA has initially issued ECHO in a "Pilot Phase" that will allow for public comment on the accuracy of facility-specific data and on the general design of the site. We take seriously EPA's repeated statements that it wants to "get its message right" about the meaning of the compliance data presented in ECHO. In that spirit, CEEI is providing these comments about the overall design of the site. Once the Agency has had an opportunity to read these comments, we hope it would be feasible to discuss the issues in greater depth with EPA.

These comments are divided into three sections. We begin with an overall summary of the primary issues that we see with ECHO. In the second section, we provide more specific comments about particular aspects of the site. Finally, we conclude with specific recommendations about steps that EPA should take in its further implementation of this project.

Overall Assessment of ECHO

In developing the ECHO Website, it appears that EPA has attempted to learn from its experience with the Sector Facility Indexing Project (SFIP) by investing more energy into pre-dissemination review to improve data quality. Specifically, EPA deserves credit for working with state agencies to reduce inaccuracies that can occur during the "hand off" from state to federal databases. CEEI also believes that the "Pilot Phase" of ECHO, with the opportunity for facilities to comment on the accuracy of the database, was a desirable step.

Unfortunately companies reviewing the records on their facilities in ECHO have been finding significant inaccuracies in the data. Some of the errors are unique to the compliance data in ECHO (e.g., did a violation occur.) More surprising, however, is the fact that companies have also been finding errors in the core facility information (e.g., facility descriptors) in the database. In light of the close coordination between EPA and the states that has occurred in the development of ECHO, we are concerned that our experience with ECHO indicates that there may be significant error rates in both the state and EPA systems.

While this accuracy issue needs to be addressed, CEEI believes that the most important set of issues raised by ECHO is how EPA is representing the state of environmental compliance and the relationship of compliance information to public risks. There is little question that ECHO is one of the most important public databases that EPA has developed in the last several years. Besides providing the new "public face" of EPA's compliance program, ECHO also represents the first time that EPA has tried to link compliance data to demographic data at the facility level for the entire country. This linkage between compliance information and demographic data necessarily draws EPA's Office of Enforcement and Compliance Assurance (OECA) into the world of risk communication.

By taking on this new objective of linking compliance data with risk-related information, EPA must also take on the responsibility of information stewardship – providing the public with

accurate and understandable information about important environmental risks. This responsibility is no longer a theoretical obligation, but rather is a clear requirement of the Information Quality Guidelines that have been issued by the Agency and the Office of Management and Budget (OMB).[3] In particular, the "objectivity" and "utility" standards in these Guidelines direct EPA to focus on the *presentation* of the information it disseminates to assure that the Agency is informing, and not misleading, its public audiences.

CEEI does not believe that ECHO, in its current iteration, is meeting those standards. The data presented in the site is often difficult to understand. In addition, because the data is primarily a measure of enforcement *activities* by EPA and state programs, it is a poor measure of actual compliance *behavior* by regulated facilities. EPA has also not yet found a way to put compliance data in a context that allows the public to understand whether companies have good compliance records. Most importantly, the link to demographic data suggests a nexus between regulatory violations and public risk that cannot be justified on the merits and is poorly articulated in ECHO at this time.

CEEI urges EPA to improve ECHO on two tracks. In the short-term, EPA should improve its explanation of the limits of the data in ECHO and provide effective opportunities for companies to clarify the record on individual facilities. In the long-term, EPA should rethink how it measures compliance and portrays the relationship of compliance to public risk.

Specific Comments

CEEI has identified four major areas of concern about the ECHO database: (1) Some of the data is inaccurate; (2) ECHO is not providing a good picture of the state of compliance; (3) Portions of the ECHO data are not well-explained; and (4) EPA has confused the public about the value of ECHO as an indicator of public risk. These concerns are explained below:

1. Data Accuracy Remains a Significant Problem

CEEI has heard reports about data inaccuracies from its members and other members of the business community that have reviewed the ECHO data on their facilities. These companies have identified several types of data errors that are arising with some regularity.[4] Some companies are reporting that they are listed for violations that they did not commit or at least were not aware of. Whether this is a question of a mistaken entry of a violation in the database or a failure of the relevant agency to notify the facility of a perceived violation is difficult to

[3] EPA, "Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility and Integrity of Information Disseminated by the Environmental Protection Agency," see http://www.epa.gov/oei/qualityguidelines/index.html, and OMB, "Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility and Integrity of Information Disseminated by Federal Agencies," 67 Fed. Reg. 8452 (February 22, 2002).

[4] In these comments, CEEI is not attempting to seek corrections of errors in particular data files. Individual companies will submit appropriate correction requests on an individual basis.

discern.[5] Either cause of the problem is a concern. Companies should not be surprised by the listing of a violation in a public database like ECHO.

Some companies are noting that they have been linked in ECHO to violations at facilities they never owned. This can only be an entry error. Companies are also reporting that company names are being linked to violations for which they are not responsible. This type of error may be an artifact of EPA's decision to use the Facility Registry System (FRS) as the system for describing facilities in ECHO.

The FRS is, in many ways, an improvement over past efforts by EPA to develop a consistent and accurate set of descriptors to identify a facility. Yet its focus is on establishing a link between multiple facility names, and thus company-specific identifiers, to a common address and geographic coordinates. FRS is not primarily designed to sort out the various ownership interests and responsible parties for particular violations. Thus companies are finding that their names are being attributed to violations at facilities that they have closed or sold to other parties. In other cases their name is linked to violations that were caused by other companies that share a common address. Thus the FRS's strength for one purpose, its ability to link multiple site "names" to a common address, is confusing the picture in ECHO of who is responsible for particular violations.

Companies are also finding errors in the facility description information about their facilities. This is troubling in light of the significant effort that EPA has undertaken to reconcile its FRS system with the facility description information in EPA program and state databases. Unfortunately these reports suggest that there may be significant errors, or at least outdated information, in the various databases from which the FRS is drawing its data.

These ongoing concerns about the accuracy of the ECHO data create much greater expectations for the error correction process that has been established for the database. CEEI applauds EPA's efforts to build a data correction mechanism into ECHO as part of its initial design. To maintain the credibility of this action, however, it will be important for EPA and the states to respond quickly to correction requests. Compliance information is one of the most sensitive forms of environmental data that EPA manages. A public database recording a violation at a facility can affect a wide range of company interests, ranging from community perception to insurance rates to a company's ability to qualify for certain business opportunities. Any errors in such information should be fixed promptly.

The universe of parties requesting corrections of data during the Pilot Phase of ECHO will, of course, be tracking EPA's responsiveness carefully.[6] CEEI is concerned, in addition, about how the correction process will respond after the Pilot Phase. Our experience with corrections in public databases like Envirofacts is that the responsiveness of the Agency varies

[5] As will be discussed later in these comments, the systems for notifying facilities about "violations" noted in inspection reports is inadequate in many instances.

[6] CEEI notes that the comments of several other parties, including the American Petroleum Institute and the American Chemistry Council, provide specific examples of some difficulties that companies are finding in using the ECHO correction tool to address inaccurate data.

quite a bit. Some correction requests are addressed in a matter of days or weeks. Other requests drag on for months. The key variable in these differences appears to be the interest and vigilance of the data steward that is responsible for the data at issue. In some states, EPA Regions and EPA programs addressing data errors is simply not a priority.

One of the generic problems with EPA's information correction systems is that there is no clear expectation on response time or a management system that makes the data stewards accountable for timely responses. A minimal component of an effective accountability system would be a mechanism for recording and publicly reporting how long it takes various data stewards to perform their tasks. Experience suggests that public measurement of performance can be an important incentive for improvements.[7]

Another important incentive for the prompt resolution of correction requests is the "flagging" system that notes in the file of a particular facility that certain data may be in error. EPA's current policy on flagging of data errors causes the flag to be posted too late in the process. A flag is not placed on a disputed data element when the correction request is filed. Instead, the flag is posted once the data steward for the information has reviewed the matter and determined that an error has occurred.

EPA's flagging policy creates the wrong incentive structure because it rewards unresponsive state and federal data stewards. If a data steward ignores a correction request, then no flag is placed on the data. This policy certainly defeats the purpose of the flagging concept from the perspective of the party filing the correction request. The policy also hides information from members of the public who would like to know the status of compliance at particular facilities. The fact that particular data about the facility is in dispute is an important element of a facility's profile that should be communicated to the public.

ECHO also highlights another aspect of EPA's flagging policy that raises concern. Currently EPA is accepting correction requests from parties besides the owners or operators of the facility. Such a step may be necessary, particularly for former owners of a facility who are trying to remove their company name from the current facility's profile. Yet EPA does not notify the current owners or operators of a facility when a third party has filed a correction request. This policy fails to recognize the central role of facility operators or owners in the management of data about their facilities. The party submitting data to EPA about its facility is the ultimate "owner" of that data.[8] EPA, and certainly any other third party who may be interested in that data, are "users" of the information.

Users of information, including EPA, do not have the right to modify the core data in a facility's file without the consent of the data submitter or at least the knowledge that such a modification is occurring. Thus there should be policies in place that provide for notification of

---

[7] Over time, this information could also then be used to develop "customer service standards" on data correction.

[8] A submitter of information to the government faces potential criminal sanctions if the data is false. Companies do not take lightly their obligations to file accurate information with government agencies. Users of information do not face the same sanctions for errors in their treatment of a facility's data.

a data submitter if other parties seek modifications of that data in ECHO or any other EPA database. By analogy, EPA has for many years provided notification to a data submitter if a third party files a Freedom of Information Act request for data that the submitter has claimed confidential.[9]

While the importance of the ECHO data warrants an upgrading of the EPA data correction process, the Agency should also be vigilant to make sure that any corrections made are sustained in the database. One of the flaws of the SFIP, which became a major frustration for data submitters, was that corrected data sometimes returned to its inaccurate state when the SFIP database was "refreshed" on a quarterly basis. While the root cause of this problem was never clearly diagnosed and explained, CEEI hopes that EPA has taken steps to avoid this problem in its design of ECHO. At a minimum, this issue should be evaluated as ECHO is implemented to assure the integrity of the data correction process.

2. ECHO Does Not Provide a Good Picture of the State of Compliance

ECHO attempts to characterize a facility's compliance record by reporting on violations of regulatory requirements in three EPA programs. The question of whether a "violation" has occurred at a facility is a linchpin of the ECHO framework. The listing of a "violation" in ECHO can be triggered by a fairly informal, even arbitrary, process. A violation may be listed in ECHO even if a facility has not received a formal Notice of Violation (NOV) from a state or federal authority. In some cases, a violation is recorded in ECHO simply because a violation is noted in a state or EPA Regional inspection report. These reports reflect, in many situations, the subjective determination of the inspector that may prove to be incorrect on further examination.

Due to the informality of the inspection process in many states, a facility may not be notified about an inspector's belief that a violation has occurred. Yet, under ECHO, that informal determination by an inspector can be the basis for identification of a violation in the database. As a result, several companies have reported that they were surprised to learn about perceived violations when they viewed the EPA ECHO profiles for their facilities. Companies should not be learning about the viewpoint of a state inspector for the first time in an EPA Website that is broadcasting that potential violation of law around the world.

This is particularly the case when EPA acknowledges that there may be systemic flaws in how violations are identified. Perhaps the best example of this concern is identified in the ECHO site itself. Under the CWA, facilities are required to file "discharge monitoring reports" (DMRs) that indicate what water discharges have occurred at a facility. If a state official fails to enter the results from a DMR into the relevant CWA database (which is used to support ECHO), then the system may assume that no DMR was submitted and that the facility has violated the law. Thus a certain percentage of CWA violations listed in ECHO reflect the failures of state and EPA staff to enter data properly, rather than any failures by the facility itself.

Given the problems that have arisen about how EPA determines that a violation has occurred, it is particularly problematic that EPA has not developed a systematic way to remove violations from the system when a mistake has been made. Several companies have identified

[9] See 40 CFR §2.204.

errors in the ECHO database and have discussed the errors with state officials. Companies are reporting conversations with state officials in which the erroneous identification of an error is acknowledged, but the state says that EPA policy prevents them from removing the violation from ECHO. State officials have indicated that EPA has no process for expunging a false determination of a violation from the ECHO system. In short, the ECHO system has an informal, low threshold system for bringing "violations" into the system, but no system for correcting the record if an error is made. CEEI does not believe that such a system is fair.

A closely related issue concerns how EPA tracks the duration of a violation in ECHO. In many cases, when a company is confronted with a violation, either by its own self-discovery or through a governmental inspection, the company moves very quickly to correct the violation. These corrective actions are often completed in a matter of days or weeks. The enforcement programs of EPA and the states, however, are much slower to react. It may take months or years to review the files on a particular compliance situation and "close out" the matter from their files. Under ECHO, the duration of a violation extends until the time that the government agrees that the violation has been corrected. As a result, the public is given the misperception that environmental violations extend well beyond the timeframe within which they are typically fixed.

This problem, and other issues arising with ECHO, is indicative of a fundamental flaw with the database. The information systems that feed data into ECHO are intended to be measures of the enforcement program at EPA. They measure the activities of the enforcing agency, not the behavior of the regulated industries. As a result, the public does not get an accurate picture of what is happening on the ground in their neighborhoods. Instead, they are receiving a picture of how various state and federal agencies are administering their programs. Thus, ECHO is a classic example of a flawed "secondary use" of an EPA database. Databases that track various measures of state and EPA program activity are being used to describe private sector behavior that may not be related.

Another aspect of the "programmatic" nature of the data underlying ECHO is the way in which it treats disputes among state and federal regulators. Companies have reported situations where state and federal regulators have different views on whether a violation of environmental requirements has occurred. Such disputes are inevitable given the complexity of modern environmental requirements. In ECHO, however, EPA has recorded only the Agency viewpoint on the existence of a violation. EPA's perspective is problematic because in most cases the applicable environmental requirements are found in state law, where deference to state officials is particularly logical.

At a minimum, EPA's unwillingness to acknowledge its dispute with state officials is misleading to the public and unfair to the data submitter. The public should be notified when a violation cited in ECHO represents a point of disagreement between the "co-regulators", state and federal, that administer environmental programs.

On a broader plane, ECHO does not help users of the database understand the overall compliance record of a facility. The database does not provide any context or point of comparison to allow the public to understand the significance of a violation. ECHO simply

indicates that a violation has occurred. It does not offer what some have called the "denominator" of a facility's compliance record – a point of comparison that allows the public to understand the facility's compliance rate and overall record.[10]

At the core of this issue is the question of what constitutes a good compliance record. Some might say that a facility is not doing a good job unless it has "no violations." Yet the reality of environmental compliance presents a more complex landscape. Most regulated facilities face EPA compliance obligations under only one Agency program. Yet there are many complex industrial sites that face a myriad of environmental responsibilities under multiple statutes, expressed as hundreds and even thousands of separate regulatory obligations.[11] In these two situations, a violation of one requirement can have very different implications. At a highly complex industrial facility with multiple permits and many separate obligations, the presence of one violation probably represents an excellent record. In a small facility that faces a few requirements, a violation might legitimately be a much greater cause of concern.

The general question of how to improve the measures of compliance behavior is a long-standing problem area for EPA. Several efforts have been made in the last several years to analyze how compliance is measured. While these efforts have been conscientious, they have not produced significant changes in how EPA and the states conduct business. ECHO, however, places the problem of measuring performance in stark relief. With ECHO the public has access to compliance data on some 800,000 regulated facilities, yet does not have a means to determine whether those facilities are doing a good job.[12]

3. The ECHO Data Are Not Clearly Explained

The data elements that constitute the building blocks of ECHO are fairly complex and are not intuitively obvious to the average user of the database. The terms used and the underlying concepts have arisen over many years as tools for EPA and state enforcement personnel to work together. As a result, users of the database who are not familiar with EPA's enforcement program can easily become confused.

As an example, the Data Dictionary for the Website lists many types of inspections. ECHO lists 25 types of Clean Air Act inspections, 13 types of Clean Water Act inspections and 14 types of Resource Conservation and Recovery Act inspections. Interestingly, the Data Dictionary lists types of inspections for other EPA programs that are not included in ECHO at this time. The program with the largest number of inspection types (51) is the Toxic Substances Control Act program.

---

[10] A good analysis of this issue is presented in a report prepared by Tischler/Kocurek for the American Chemistry Council and the American Petroleum Institute entitled "Compliance Rate Denominator Study" (January 2002).

[11] As an example, see report by Horizon Environmental Corporation entitled "Final Environmental Regulatory Profile" (September 25, 1998) prepared for the American Automobile Manufacturers Association.

[12] There are at least two dimensions for expressing the significance of a violation. First, what is the total universe of regulatory obligations that a facility faces? Second, how much risk is associated with a particular violation? The latter question will be discussed later in the discussion of "significant non-compliance."

Besides the sheer number of inspections listed, the definitions of the various inspection types introduce concepts that are important to the administration of federal and state programs, but will not be meaningful to most ECHO users. For example, OECA has a management tool called the Reporting for Enforcement and Compliance Assurance Priorities (RECAP) that measures whether the states are investing their inspection resources in areas that are important to EPA CAA enforcement personnel. In the Data Dictionary, EPA has articulated a coding system that distinguishes inspections that "count" under the RECAP system and those that do not. For the user of the ECHO database, this coding system is difficult to follow and certainly impossible to interpret. The public simply cannot determine the significance of the labels attached to the various inspection types.

Another area of confusion in the ECHO nomenclature arises when the database refers to a violation as being "addressed" or "unaddressed".[13] A normal understanding of those terms would suggest that an "unaddressed" violation means that the facility is continuing to violate the applicable requirements. Yet the term actually refers to whether the relevant state or EPA regulatory authority has made a decision at the facility. Thus, a company could quickly correct a violation at its facility within a few days after its discovery and that violation could be listed as "unaddressed" in the state and EPA databases for months and even years awaiting the attention of state or EPA enforcement personnel to close the file and shift the violation from "unaddressed" to "addressed" status.

This is another example of the anomalies that arise when a measure designed to track the EPA enforcement program is used as a surrogate for measuring the compliance behavior of regulated facilities. The explanations of the data offered in ECHO do not make this distinction clear. As a result, data submitters are frustrated by the apparent unfairness of the system, and the public is not receiving a clear picture of the actual state of compliance.

EPA has made some effort to identify and explain the data limitations of ECHO. Much of that explanation is provided in the Data Dictionary for the site. EPA has also compiled a listing of "Known Data Problems" that can be accessed from the opening page of the Website. While EPA deserves credit for attempting to explain ECHO's limits, EPA needs to make these explanations clearer. The Data Dictionary is a very dense document that requires a good understanding of federal environmental programs, EPA operations and state-federal relationships in order to interpret and understand what is being communicated. The document needs editing to make it accessible to a more general audience.

The information on known data problems should also not be set off in a separate section that is not linked to the facility reports. A user of the Website is unlikely to review the entire section on Known Data Problems. Users would like to know, however, when a known data problem may be influencing particular facility reports. For example, EPA has noted that there may be problems with water violations related to chlorine in Virginia. If such a violation appears at a Virginia facility, users would like to be made aware of that issue when viewing reports on Virginia facilities.

[13] This language appears, for example, in the coding system for CAA High Priority Violations.

4. EPA Has Confused the Public about ECHO's Value as an Indicator of Public Risk

Perhaps the most important decision made in the design of ECHO was the effort to align facility-specific compliance data with the demographics of the community around the facility. This step was certainly not necessary because compliance information is not inherently related to immediate community risks. In many ways, it is puzzling that EPA has decided to present community demographic data in a compliance database, as opposed to a "background" database like the FRS or a database that reports on environmental releases.

Since inclusion of demographic data in a compliance database is not an obvious step, all parties viewing ECHO are assuming that the Agency is attempting to communicate some type of message about public risks with ECHO. The decision to align demographics and compliance data can have no other purpose. If ECHO remains in its current form, OECA is obligated to fulfill the risk communication responsibilities it will have assumed. At a minimum, OECA must assess how it is satisfying its obligations under the "objectivity" and "utility" standards in the Agency and OMB Information Quality Guidelines.[14]

The core problem in using the compliance data underlying ECHO as a statement about public risk is that the violations highlighted in the database, even assuming they are accurate, may not have any direct relationship to community risk. ECHO reports on what EPA calls "significant non-compliance" (SNC), a term of art used in EPA's enforcement programs.[15] The three programs that are included in ECHO have developed their own definitions of this term over time as they have worked with the states to develop programmatic priorities. The three programs do not share a common definition of the term.

The concept of an SNC was not developed as a means to articulate the risks of industrial facilities to communities. EPA uses the SNC label as a means of articulating the priorities of the enforcement program. Certainly public risk is a factor that is considered when deciding that a particular violation will be given the SNC label. Yet public risk is not the only factor. Some violations receive the SNC label because they are important for *program administration* of particular regulations, even though there is no immediate risk associated with these violations.

Several examples make the point. Under the CAA, EPA has given priority status to the failure to submit a Title V certification. Certain other testing, monitoring and recordkeeping obligations are also high priority violations under the CAA enforcement program. These paperwork violations do not present immediate public risks, but EPA certainly could conclude that they are essential components in administration of the program and warrant attention in their compliance program.

---

[14] The linking of databases, with an apparent intent to communicate with the public about risk, may also trigger the obligations under the Agency and OMB Guidelines applicable to "analysis of risks to human health, safety and the environment." These obligations include a consideration of the principles articulated in the Safe Drinking Water Act for the development of scientific analysis and presentation of that analysis to the public.

[15] CEEI recognizes that under the CAA program EPA uses the term "High Priority Violation" to identify priority violations. Since both the CWA and RCRA programs use the term "Signficant Non-Compliance" for the same purpose, CEEI will use that term generically to capture the concept for all three programs.

Both the CWA and the RCRA program are similarly dependent on compliance with certain paperwork obligations. Under the CWA, for example, the filing of DMRs is a critical program element that EPA has identified as a priority. All of the programs identify repeat violations of the same provision as a priority matter. It is understandable that repeat violations could be of concern to a compliance program, even though the underlying violations involve paperwork requirements.

CEEI's concern with the SNC concept is not a disagreement over whether the differing approaches to priority-setting found in the three programs make sense as a *means of administering those programs.* Our concern is that these statements of program priorities are not *reliable statements about the day-to-day risks* that people living in a community are experiencing from a nearby facility. ECHO could be telling a community that they live within one mile of a substantial release of air toxics or of a missing RCRA training plan. The underlying data found in ECHO cannot differentiate or communicate the distinction between these two, very different situations. EPA should not be attempting to communicate risks to the public with data that was never intended for this purpose and simply cannot credibly be used as a surrogate for risk.

Our general concern about the use of compliance data to communicate risks to the public has been heightened by the current explanation of the demographic data in ECHO. As stated above, the concept of a SNC does not have consistent definition across EPA, but rather is grounded in the unique policies of each EPA program. For the sophisticated users of this database, it is important to have access to the core program documents that explain the SNC definitions.

The Data Dictionary for ECHO identified several key guidance documents that explain the SNC framework:

- For CAA: EPA Memorandum entitled "Issuance of Policy on Timely and Appropriate Enforcement Response to High Priority Violations" (December 22, 1998);
- For CWA: "A General Design for the SNC Redefinition Enhancement in PCS" (December 12, 1996); and
- For RCRA: EPA Memorandum entitled "A Hazardous Waste Civil Enforcement Response Policy" (March 15, 1996) and EPA Memorandum entitled "A Transmittal of Addendum to the 1996 Hazardous Waste Enforcement Response Policy" (April 25, 2000).

In an effort to review these documents, we searched the EPA Website for the documents in January 2003. We were only able to find the CAA guidance memorandum on-line. We then contacted OECA staff to find the remaining documents. They were able to supply us with the RCRA memoranda. They referred us to Office of Water (OW) staff for the CWA document. Despite several contacts with the OW staff, we never received the document from EPA. While the EPA staff we contacted on this matter tried to be helpful, the exercise indicates that the underlying documents defining what is an SNC are not readily available, even to EPA and state enforcement staff, much less for members of the public who are persistent and more knowledgeable about EPA operations.

More broadly, the explanations that EPA offers of the demographic data in ECHO are confusing. The introductory text to the "Demographic Profile" section found in each facility profile states the following:

> This section is to provide context regarding the community setting of the facility. No relationship between this information, and the other data included in this report is implied.

The Data Dictionary elaborates on this language:

> The total population within a given radius of a facility is one indicator of the facility's surrounding environment and provides context for other facility indicators...This indicator does not imply that there is any exposure to the identified population. Data are for information purposes only and do not indicate any association with other sections of this report.

EPA's intent with these statements is almost incomprehensible. The demographic data is seen as providing "context", yet there is "no relationship" between that data and the other elements of the ECHO report. If there is no relationship between the demographic data and the rest of the ECHO report, how can the demographic data be providing relevant "context" for the public? EPA says that the demographic data "does not imply that there is any exposure to the identified population" yet the data is offered for "information purposes." Presumably any information product is issued for "information purposes", so this statement only begs the question of what purpose EPA believes should be served by the demographic data.

CEEI does not view the problem with the demographic data to be a matter of artful crafting of a disclaimer statement. The public has a legitimate right to believe that EPA included the demographic data in the ECHO database for a purpose. It is entirely reasonable for the public to presume that EPA's decision to link the demographic data to a compliance report using terms like "significant non-compliance" was based on a belief that there was a direct relationship between the reported violations and public risk. The public should not have to wade through a series of contradictory EPA statements that are buried in highly technical "data dictionaries" to determine that EPA may not think such a cause and effect relationship exists.

More fundamentally, if EPA believes that the public should draw no implication about the potential for public risk from the compliance data as presented in ECHO, then the Agency should not be taking any actions that encourages such a misperception. In its current state, EPA's disclaimer about the demographic data is worse than neutral. The disclaimer is so confusing that it becomes misleading to the public and is clearly inconsistent with the "utility" and "objectivity" standards in the Information Quality Guidelines.

Recommendations

CEEI believes that EPA should undertake several measures to improve the ECHO information product. The list of recommendations provided below begins with actions that

should be a focus of attention in the short-term and then ends with longer term measures that should be considered:

- The error correction measures that are under way during the Pilot Phase of ECHO should be continued after the Pilot Phase is over. The error correction process should be made more timely on a uniform basis among all data stewards. To facilitate that process, EPA should begin publishing public statistics on how the various data stewards are responding to correction requests. These statistics need not be limited to ECHO data. This reporting function should mature into the setting of "customer service standards" for timely corrections over time.

- EPA should evaluate whether previously corrected data is returning to its "error" state once the ECHO data is refreshed. If this phenomenon is occurring, EPA and the states should undertake a "root cause" assessment of the problem and make appropriate software or management changes to eliminate the problem.

- EPA should change its approach to notification once a data correction request is filed. The data that is the subject of a correction request should be flagged in the ECHO site once the correction request is filed. In addition, a data submitter should be provided with direct notification (potentially as an email) if a third party files a correction request related to the submitter's data.

- Before the posting of a violation in ECHO, EPA should make sure that the affected facility has been notified of the violation. If the facility has not previously received a formal notification from state or EPA Regional officials, then OECA should not post the violation.

- If there is a dispute between EPA and a state about the presence of a violation at a particular facility, that dispute should be resolved before a violation is posted in the ECHO Website. If EPA insists on posting its view of the compliance status of a facility, the Agency should post the fact that the relevant state disagrees with the Agency's view.

- EPA should establish a procedure for expunging a violation from the ECHO database if it is determined that an initial finding of a violation was in error. Mistaken violations should not be carried as a violation that is subsequently corrected in the ECHO database.

- EPA should remove from the ECHO database violations that may be triggered by systemic data management errors, such as the automatic entry of a violation if a DMR is not entered in agency databases.

- In describing the duration of a violation, EPA should reflect how long the violation actually continued and not how long it takes the relevant state or EPA enforcement staff to review the situation and close its files. If that result cannot be achieved with existing data, then EPA should not attempt to describe the duration of a violation in ECHO at this time.

- EPA should improve the design of the ECHO Website and the explanations of the data to make the limitations of the data clearer. As a starting point, EPA should systematically "beta-test" this Website with members of the general public to determine how they use the site (e.g., what sections are read and what sections are ignored) and what conclusions they draw from the current information. EPA can then develop and test with consumers descriptions of the information that give the public a better picture of what questions the data can and cannot answer.

- Information about known data problems, which are now explained in a general section of the Website, should be linked directly to the specific facility reports that might be implicated by the general problem. This step assures that the information about known problems remains in the mainstream of the ECHO site.

- Despite the efforts described above, there will be aspects of the compliance picture at individual facilities that cannot be conveyed accurately because of the complexity of the information and information that is not available in EPA's compliance databases. CEEI recommends that EPA create a field in the report on each facility that allows a facility owner or operator to provide key information that is otherwise missing from the report. The facility owner or operator could place in this field a brief textual message or a URL address for a more fulsome description of the facility and its compliance record.[16]

- EPA should reverse its decision to link compliance data and demographic data in the ECHO Website. Given the fact that the current EPA definitions of an SNC are not necessarily risk-based and the clear difficulty EPA is having in developing understandable disclaimers, EPA should remove the demographic data from ECHO for the time being.

- If EPA intends to develop a public Website that identifies violations that have a direct impact on health and the environment, the Agency will need to develop an alternative to its current concept of "significant non-compliance" as the linchpin for such a system.

- Companies have identified significant errors in how their facilities are described, which suggest that there may be important accuracy problems with the FRS. While CEEI supports the idea of a common facility identification system and is willing to support the FRS as the platform for such a system, the facility identification system must be highly accurate. EPA, the states and the universe of facility managers need to find a more effective way to improve the accuracy of the FRS. A system that relies on EPA's issuance of controversial Websites, such as ECHO, in anticipation of a wave of correction requests is not an optimal way to improve the accuracy of the Agency's facility identification system.

- If EPA wants to continue to use FRS as the linchpin for the ECHO system, the Agency will need to examine how it can link the two systems so that the parties currently

[16] EPA has already established a precedent for such a step. In the Toxic Release Inventory component of Envirofacts, EPA has for several years established a link to each facility's profile in the Environmental Defense Scorecard Website.

responsible for compliance at a facility are more clearly identified. EPA should avoid the situation where a search of a corporate name identifies a set of facilities that the company has closed or no longer owns.

- EPA should reconsider how it measures and describes the compliance record of regulated facilities. ECHO underscores the fact that the information about a violation at a facility is not being placed in context, either in terms of the number of requirements facing particular facilities or of the environmental impact of the violation. Over the years there have been several pilot projects that have developed interesting models for explaining compliance information to the public. As an example, the automobile industry has worked with the State of Michigan to develop the Community Environmental Awareness Project, an effort to provide the public with a more integrated profile of how automobile assembly plants are performing under environmental laws. CEEI urges EPA to review recent experience with such innovative efforts and develop a strategy for improving measurements of compliance.

Conclusion

The ECHO information product is EPA's first systematic effort to present compliance data at a national level for over 800,000 facilities. It is an ambitious effort. EPA has certainly made a serious effort to learn from its past experience with public databases on enforcement and compliance information. It is not surprising, however, that a Website of this scope covering the inherently controversial subject of compliance would need further refinement.

CEEI believes that EPA should focus it energies on the basic accuracy of the data in the underlying databases that support ECHO, making sure that the Agency can make appropriate corrections on a timely basis. EPA should also avoid using the available compliance data to answer questions that cannot be fairly answered. This means that EPA will need to scale back some of the compliance metrics that it is trying to present in ECHO right now, such as the duration of a violation. In addition, EPA should focus on portraying the compliance record of regulated facilities, without attempting to draw any conclusions about how compliance records are affecting public risk.

Over time, EPA may decide to refine the information it collects and portrays in public databases on compliance. Those efforts, however, should begin with a more strategic vision of what questions EPA hopes to answer, a plan to assemble the right data and a focus on "getting the message right" so that the public has access to data that is understandable and can provide high quality answers to their questions.

Coalition for Effective Environmental Information
March 31, 2003

7169871.2